

УДК 621.3

**МУЛЬТИФРАКТАЛЬНАЯ СТРУКТУРА ШЕПОТА  
И РАСПОЗНАВАНИЕ РЕЧЕВЫХ СТРУКТУР****канд. техн. наук, доц. В.И. СОЛОВЬЕВ***(Восточноукраинский национальный университет им. В. Даля, Луганск);***д-р техн. наук, проф. О.В. РЫБАЛЬСКИЙ***(Национальная академия внутренних дел Украины, Киев);***д-р техн. наук, проф. В.К. ЖЕЛЕЗНЯК***(Полоцкий государственный университет)*

*Демонстрируются результаты исследований по выявлению мультифрактальных структур для гласных фонем шепота. Экспериментальные исследования по частотно-временной модификации гласных фонем русского языка, произносимых шепотом, показали их явно выраженную мультифрактальную структуру. Выявлена возможность «конструирования» гласных фонем шепота из «атомарных» структур фонем путем аффинных преобразований.*

Исследования и практические разработки в области автоматического распознавания речи в последнее время базируются на многообразных подходах. В ранних периодах основой этих исследований являлись вероятностные модели скрытых марковских цепей. Эти модели широко применялись во второй половине прошлого столетия для решения задач в области речевых технологий [1–3]. Сегодня в исследованиях и разработках для распознавания речи используются идеи нейронных сетей.

В настоящее время созданы практические системы этого типа, доказавшие свою эффективность в устройствах мобильной связи. Такие системы быстро совершенствуются. Детали и технологические тонкости существующих систем не публикуются в открытой печати, поэтому оценка эффективности подобного рода систем затруднительна. С другой стороны, с времен Гельмгольца [4–6] и до настоящего времени отсутствуют явно выраженные теоретические основы распознавания речи путем анализа фрагментов сигналов звуковых акустических волн. Несмотря на огромное число разнообразных исследований в этой области, отсутствуют последовательные теоретические концепции, обеспечивающие возможность эффективного распознавания речи. Кроме того, практически отсутствуют исследования по распознаванию речи, генерируемой в форме шепота. Характеристики речи шепотом существенно отличаются от характеристик обычной речи, тем не менее человек однозначно распознает шепот.

Полагая, что научно обоснованные модели распознавания речи, в конечном счете, должны обеспечить автоматическое распознавание речи, генерируемой в виде акустических колебаний любым известным способом, в том числе и шепотом.

С другой точки зрения, шепот может представлять собой, в определенном смысле, более простую физическую модель для построения эффективной системы автоматического распознавания речи. Принято считать, что в шепоте отсутствуют характеристики индивидуальных особенностей голоса, оказывающих существенное влияние на процесс распознавания обычной речи.

В данной работе представлены результаты исследований по распознаванию гласных фонем при генерации речи шепотом.

**Постановка задачи исследования.** Задачу распознавания проанализируем контекстно независимой речи с неограниченным языковым словарем, генерируемой шепотом, исследуя её на уровне фонемических составляющих русской речи. При этом в исследовании рассматриваются только гласные фонемы.

Рассмотрим, например, фрагмент речи (фонема [а]) в виде графика звуковой волны, сгенерированной шепотом, представленный на рисунке 1 (здесь и далее для построения графиков и иллюстраций используется программный комплекс «Фрактал» [7]).

Вид графика звуковой волны и ее спектра отличается при различном темпе речи, конкретном контексте так, как и другие параметры фрагментов звуковой волны.

Важным фактором приводимых в дальнейшем результатов разработки и исследований являются критерии соответствия разрабатываемой модели объективно существующей природе кодирования и декодирования речевой информации. В качестве главного методического направления решения этого вопроса принята следующая методика исследования.

Фрагменты звуковой волны, эквивалентные при прослушивании фонемам русского языка, целенаправленно модифицируются в соответствии с разрабатываемой физической моделью (во временной и в частотной областях). После модификаций эти фрагменты прослушиваются группой испытуемых с целью выявления идентичности фонемического звучания первичного фрагмента и модифицированного. Иден-

точность распознавания фонемы до и после модификации определяется на группе испытуемых в случае 95 % распознавания фонемы.

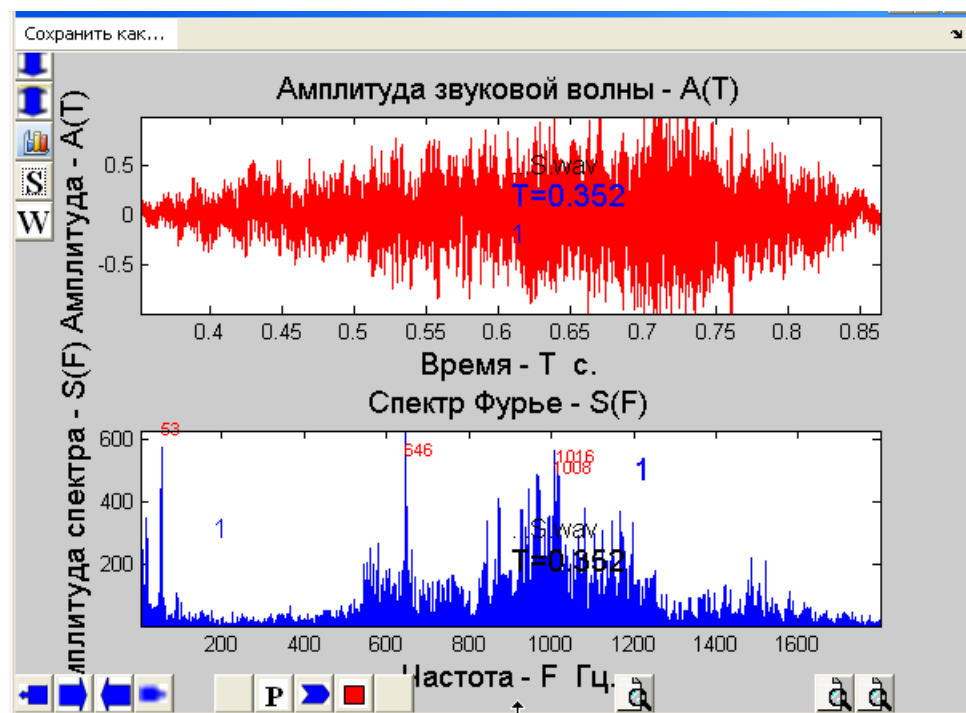


Рис. 1. Фрагмент речи (шепот, фонема [a])

Экспериментально исследовались отдельно произносимые шепотом гласные фонемы с фонемами в рамках различных текстов и характеристик шепота.

Поставим задачу выявления из гласных фонем, генерируемых шепотом, “атомарных” структур на интервалах времени 5...15 мс. Эти структуры должны полностью характеризовать конкретную гласную фонему и позволить воссоздать основные характеристики шепота по выделенной “атомарной” структуре.

**Мультифракталы и устойчивые фонемические структуры шепота.** Рассмотрим фрагменты речи в аудиоданных шепота как дискретный временной ряд амплитуды звуковой волны, а также задачу выявления самоподобных структур в речевом фрагменте временного ряда, на основе которых возможно формирование фонем. Самоподобие в дальнейшем далее рассматриваем как самоподобие мультифрактальных структур в соответствии с концепциями Мандельброта [8–13]. Это примерное геометрическое подобие визуально наблюдаемых фрагментов амплитуды звуковой волны, как и в представленном на рисунке 2 фрагменте звуковой волны, эквивалентном фонеме [a], произнесенной шепотом. В нижней части рисунка 2 – спектр Фурье выделенного фрагмента фонемы. Как известно, спектры отдельно выделенных определенных гласных звуков обычной речи имеют некоторое геометрическое подобие между собой [5; 6]. На рисунке 2 видно, что какие-то элементы геометрического подобия присутствуют и на фрагментах шепота, однако они слабо выражены при визуальном рассмотрении.

Рассмотрим самоподобие фрагментов амплитуды звуковой волны фонемы [a] на малых интервалах времени (порядка 5...15 мс). Для этого выделим фрагмент этой фонемы в области  $T = 0,55...0,555$  с, показанный на рисунке 3, и назовем его “атомарным” фрагментом.

Сигнал “атомарного” фрагмента во временном представлении (см. рис. 3) начинается с нулевого отсчета. Спектр данного малого фрагмента уже существенно отличается от спектров с большими временными интервалами для фонемы [a]. Также он отличается и по частотному разрешению ввиду уменьшения временного интервала.

Поставим задачу “конструирования” фонемы [a] из этого малого “атомарного” фрагмента фонемы путем аффинных преобразований. Составим из фрагмента, представленного на рисунке 3, модифицированный акустический фрагмент из нескольких десятков полностью аналогичных фрагментов, отличающихся друг от друга растяжением-сжатием по амплитуде, как это показано на рисунке 4.

При прослушивании модифицированного фрагмента испытуемыми четко идентифицируется фонема [a]. При этом появляются характеристики голоса (произвольные) с частотой основного тона, однозначно зависящей от длительности выбранной для модификации “атомарной” структуры.

Как показывают многочисленные эксперименты, при различных характеристиках шепота при прослушивании совершенно четко идентифицируется искусственно сконструированная фонема [a]. При этом индивидуальные характеристики голоса при подобном конструировании зависят как от длительности “атомарной” структуры, используемой для модификации, так и от характера аффинных преобразований.

Появление характеристик голоса обусловлено косвенным привнесением в модифицированный фрагмент частоты основного тона, зависящей от выбранной длительности фрагмента шепота.

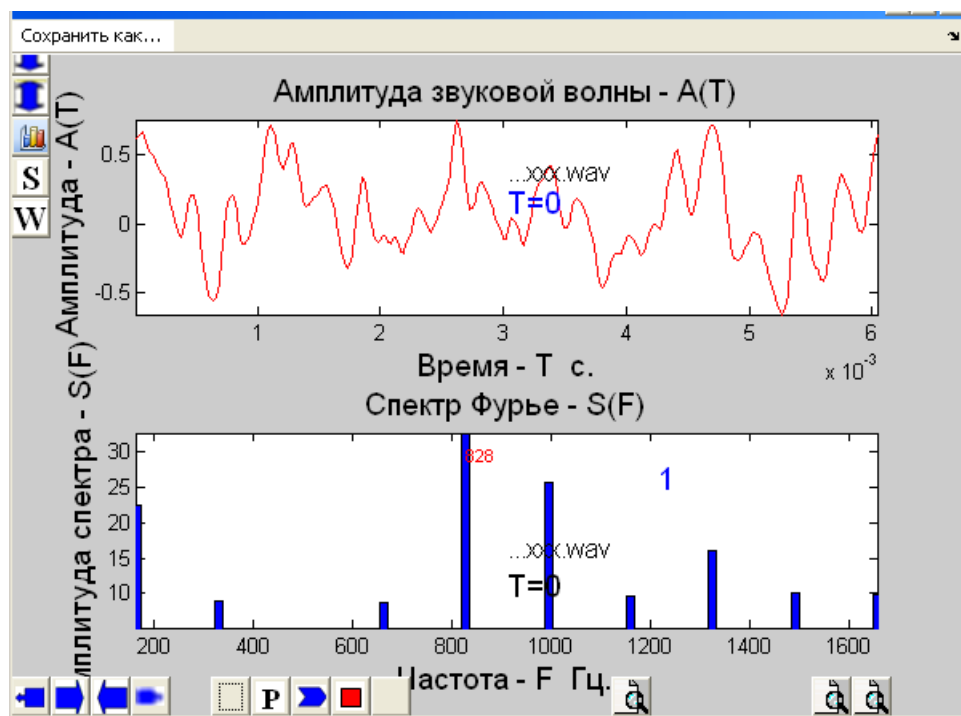


Рис. 2. Фрагмент акустической волны (шепот, фонема [a])

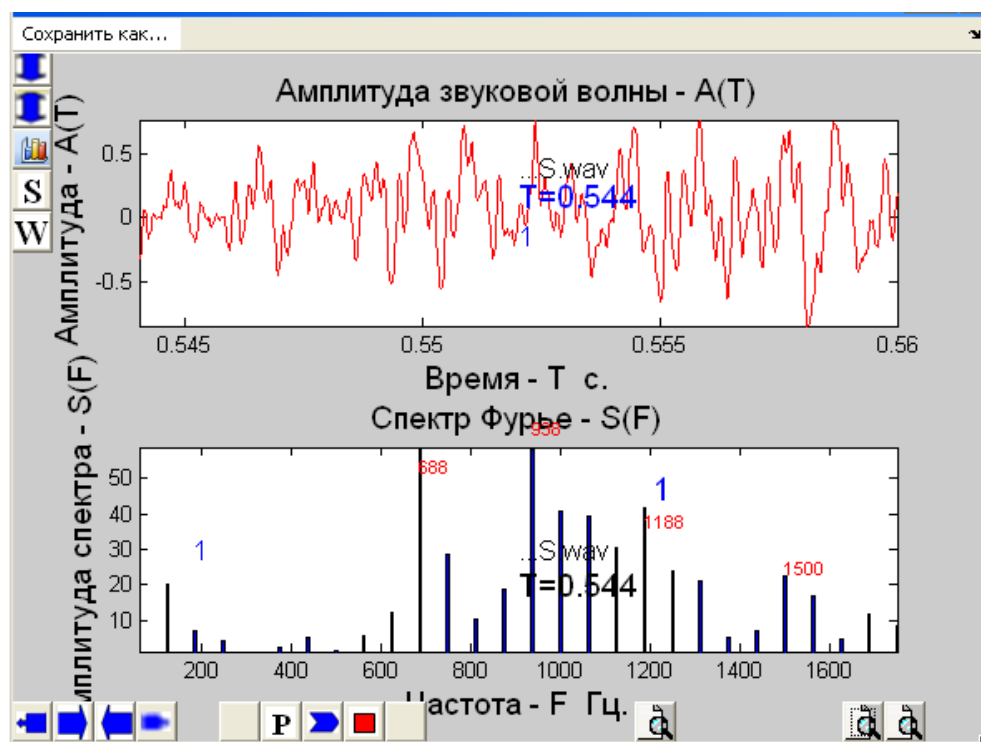


Рис. 3. “Атомарный” фрагмент фонемы (шепот, [a])

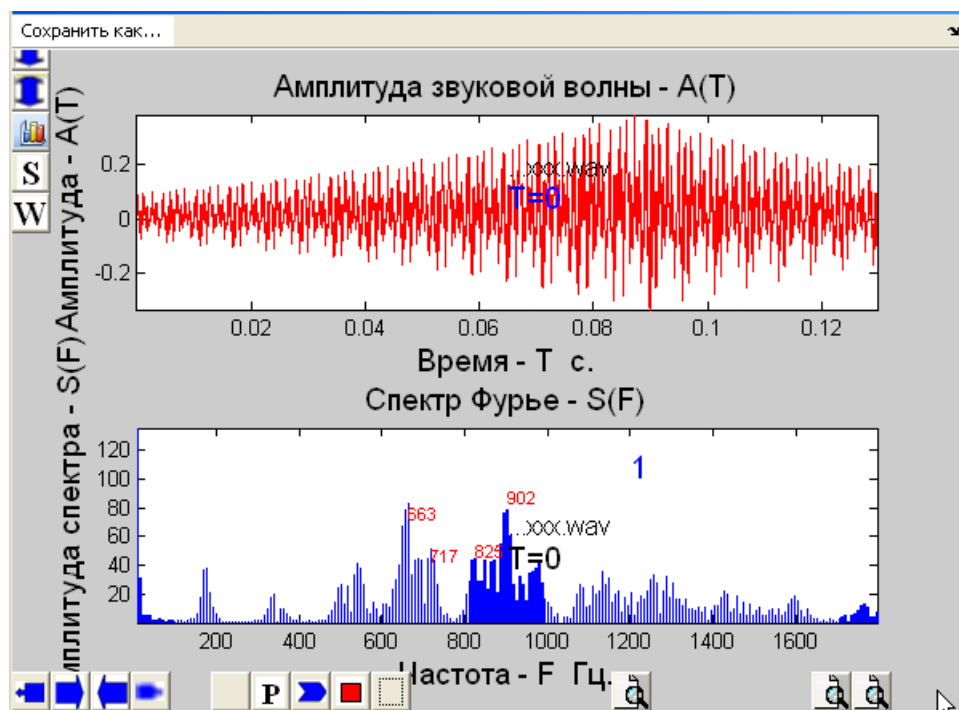


Рис. 4. Фрагмент модифицированной звуковой волны (шепот-фонема [а])

Необходимо отметить, что при выборе “атомарной” структуры шепота не любой выбор дает строго идентифицируемую при дальнейших модификациях фонему [а]. Как показывает анализ, это связано с наличием в фрагментах шепота фрагментарно появляющихся нестационарных участков, не содержащих информации о фонеме [а]. Но, примерно 70...80 % “атомарных” фрагментов дают при модификациях достаточно строгую идентификацию.

Приведенные выше результаты исследований показывают, что гласная фонема [а], произнесенная шепотом, может быть представлена (“сконструирована”) на основе определенных “атомарных” фрагментов этой же фонемы путем серии аффинных преобразований растяжения-сжатия по амплитуде “атомарного” фрагмента. Это является экспериментальным подтверждением наличия мультифрактального подобию речевых фрагментов амплитуды звуковой волны, наблюдаемого визуально.

Аналогичная серия экспериментов показала правомочность описанного выше мультифрактального представления для произносимой шепотом фонемы [и]. (Гласные фонемы рассматриваются согласно Московской фонологической школе).

Другие гласные фонемы русского языка – [о], [у], [э] также имеют мультифрактальную структуру, позволяющую осуществить “конструирование” фонемы из “атомарных” составляющих путем их аффинных преобразований. Однако для этих гласных мультифрактальные структуры являются более сложными объектами. Мультифрактальные модели этих гласных фонем (для шепота) требуют дальнейших исследований.

**Заключение.** Экспериментальные исследования по частотно-временной модификации гласных фонем русского языка, произносимых шепотом, показали их явно выраженную мультифрактальную структуру. Показана возможность “конструирования” гласных фонем из “атомарных” фрагментов фонемы путем аффинных преобразований.

## ЛИТЕРАТУРА

1. Марков, А.А. Об одном применении статистического метода / А.А. Марков // Изв. АН. – 1916. – Сер. 6, X. – № 4, 239.
2. Elinek, F. Распознавание непрерывной речи статистическими методами / F. Elinek // ТИИЭР 64. – 1976. – № 4. – С. 131–160.
3. Elinek, F. Разработка экспериментального устройства, распознающего раздельно произнесенные слова // ТИИЭР 73. – 1985. – № 11. – С. 91–99.
4. Von Helmholtz, H. Die Lehe von Tonempfindungen / H. von Helmholtz. – Brannschweig, Vieweg, 1863.

5. Фланаган, Дж. Анализ, синтез и восприятие речи / Дж. Фланаган; пер. с англ. / под ред. А.А. Пирогова. – М.: Связь, 1968. – 396 с.
6. Фант, Г. Анализ и синтез речи / Гуннар Фант; пер. с англ. В.С. Лозовского.
7. Рыбальский, О.В. Система идентификации аппаратуры аудиозаписи на основе мультифрактального подхода / О.В. Рыбальский, В.И. Соловьев // Вісник Східноукраїнського нац. ун-ту. – 2010. – № 9 (151). – С. 58–64.
8. Mandelbrot, B. Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis / B. Mandelbrot // Annals of Economic Social Measurement 1, 1972.
9. Mandelbrot, B. The Fractal Geometry of Nature / B. Mandelbrot. – New York: W. H. Freeman, 1982.
10. Mandelbrot, B. A Multifractal Walk Down Wall Street / B. Mandelbrot // Scientific American, 1999.
11. Mandelbrot, B. Robustness of the rescaled range R/S in the measurement of non-cycling long-run statistical dependence / B. Mandelbrot // Water Resources Research. – 1969. – № 5. – P. 967–988.
12. Павлов, А.Н. Мультифрактальный анализ сложных сигналов / А.Н. Павлов, В.С. Анищенко // Успехи физических наук. – 2007. – Т. 177, № 8.
13. Федер, Е. Фракталы / Е. Федер. – М.: Мир, 1991. – 326 с.
14. Психоакустические аспекты восприятия речи. Механизмы деятельности мозга / под ред. Н.П. Бехтеревой. – М.: Наука, 1988. – 504 с.
15. Цвикер, Э. Ухо как приемник информации / Э. Цвикер, Р. Фельдкеллер; пер. с нем. под ред. Б.Г. Белкина. – М.: Связь, 1971. – 225 с.
16. Алдошина, И.А. Основы психоакустики. Звукорежиссер / И.А. Алдошина. – 2000. – № 6. – С. 36–40.
17. Сорокин, В.Н. Теория речеобразования / В.Н. Сорокин. – М.: Радио и связь, 1985. – 312 с.
18. Малла, С. Вейвлеты в обработке сигналов / С. Малла. – М.: Мир, 2005. – 670 с.

Поступила 08.09.2014

#### MULTIFRAKTAL STRUCTURE OF WHISPER AND RECOGNITION OF SPEECH STRUCTURES

**V. SOLOVYOV, O. RYBALSKY, V. ZHELEZNYAK**

*The article describes the results of the research to identify multifractal structures for transparent phonex whisper-quiet performance. Shows ability to “design” transparent phonex whisper-quiet performance from the “atomic” structures consonantal clusters by an affine transformation.*